TRANSPORT
ISSN 1648-4142 print / ISSN 1648-3480 online
2012 Volume 27(2): 149–157
doi:10.3846/16484142.2012.692354

DEFINING LOS CRITERIA OF URBAN STREETS USING GPS DATA: *k*-MEANS AND *k*-MEDOID CLUSTERING IN INDIAN CONTEXT

Prasanta Kumar Bhuyan[1], Kurra Venkata Krishna Rao[2]

[1]*Dept of Civil Engineering, National Institute of Technology, 769008 Rourkela, India*
[2]*Dept of Civil Engineering, Indian Institute of Technology Bombay, 400076 Mumbai, India*
E-mails: [1]pkbtrans@gmail.com (corresponding author); [2]kvkrao@civil.iitb.ac.in

Submitted 18 August 2010; accepted 3 November 2010

Abstract. The objective of this study is to classify urban streets into a number of classes and to define the speed ranges of levels of service (LOS) categories in Indian context. In this purpose, average travel speed on street segments is used as the measure of effectiveness, which has been obtained from second-wise speed data collected using Global Positioning System (GPS) receiver. Midsized vehicle (car) was used to collect travel speed data on five urban road corridors comprising of 100 street segments in the city of Mumbai and two major road corridors of Kolkata city in India. Both *k*-means and *k*-medoid clustering methods and several cluster validation measures have been employed in the classification of urban streets and LOS categories. It is found that *k*-medoid clustering is more suitable in Indian context and speed ranges of level of service categories are significantly different from that values mentioned in HCM 2000.

Keywords: level of service, urban streets, GPS, clustering, cluster validation measures.

1. Introduction

A level of service is not well defined for highly heterogeneous traffic flow on urban corridors in India, in this regard; an attempt has been taken to define the level of service criteria in this study. Urban Street LOS normally based on average through-vehicle travel speed for the segment or for the entire street is under consideration. The use of moving observer method is the most commonly used technique for collecting travel time data. However, accuracy with this technique varies from technician to technician. Recent research has demonstrated the feasibility of using Global Positioning System (GPS) receiver in recording location as latitude-longitude, travel time and travel speed.

Level of service in the Highway Capacity Manual (2000) is defined as 'a quality measure describing operational conditions within a traffic stream, generally in terms of service measures such as speed and travel time, freedom to maneuver, traffic interruptions, comfort and convenience'. The Highway Capacity Manual (2000) also designates six levels of service for each type of facility, from 'A' to 'F' with LOS 'A' representing the best operating conditions and LOS 'F' the worst. As in Highway Capacity Manual (2000), IRC 106… (1990) guideline has been suggested that on urban roads, the LOS is affected strongly by factors like the heterogeneity of traffic, speed regulations, frequency of intersections, presence of bus stops, on-street parking, roadside commercial activities, pedestrian volumes etc. For heterogeneous traffic condition in India, Marwah and Singh (2000) LOS has been classified into four groups (I–IV). Similarly, Maitra *et al.* (1999) redefined the LOS boundaries into nine groups 'A' to 'I' by quantifying congestion as measure of effectiveness for heterogeneous traffic flow condition.

Defining level of service criteria it is basically a classification problem and the cluster analysis is the most suitable technique that can be applied for the solution to it. Prassas *et al.* (1996) applied the cluster analysis tools to a set of traffic engineering data in which deterministic modeling and regression analysis had been applied. Lingras (1995) compared grouping of traffic in the classification of traffic patterns using the Hierarchical Agglomerative Clustering and the Kohonen Neural Network methods. Lingras (2001) applied Hierarchical Agglomerative Clustering technique and an evolutionary Genetic Algorithms approach for the classification of highway sections. Oh and Ritchie (2002) used *k*-means, Fuzzy and Self Organizing Map (SOM) clustering in a real-time signalized intersection surveillance system for the classification of LOS categories. Kim and Yamashita (2007) used *k*-means clustering algorithm to examine patterns of pedestrian crashes in Honolulu, Hawaii.

The data set that was used in this study was obtained from 10÷12 travel runs taken on five major urban corridors in the city of Mumbai. The total length of these corridors is about 140 km. These corridors, on the whole, were divided into 100 street segments. *k*-means and *k*-medoid clustering techniques are employed to define speed ranges of LOS categories of urban street classes. The algorithms used in this study are based on the search of *k* representative objects called means and medoids of the clusters. After finding a set of *k* representative objects, *k* clusters are constructed by assigning each object of the data set to the nearest representative objects. The methodology of applying cluster analysis and various validation parameters and their use as a measure of optimum clustering has been demonstrated. Average free-flow speeds on street segments were classified into four groups to get free-flow speed ranges of urban street classes. Next, average travel speeds (collected during peak and off-peak hours) on all segments under each street class were used as input data for clustering to get the speed ranges for each level of service categories A to F. In order to justify the applicability of this level of service criteria in Indian context, similar set of data were collected from Kolkata City and validated the Speed ranges of levels of service criteria.

2. *k*-means and *k*-medoid Clustering

One of the most well-known hard partitioning methods is *k*-means clustering. It is most useful for forming a small number of groups from a large number of observations. It requires variables that are continuous with no outliers. The function *k*-means partitions the observed data into *k* mutually exclusive clusters, and returns a vector of indices indicating to which of the *k* clusters it has assigned at each observation. *k*-means uses an iterative algorithm that minimizes the sum of distances from each object to its cluster centroid. This algorithm moves objects between clusters until the sum cannot be decreased further. However, *k*-means does not guarantee unique clustering because we get different results with randomly chosen initial clusters. The *k*-means algorithm gives better results only when the initial partitions are close to the final solutions (Jain, Dubes 1988).

Kaufman and Rousseeuw (1990) presented a *k*-medoid algorithm which they call PAM (Partition Around Medoids). This algorithm attempts to minimize the total distance between objects within each cluster. The algorithm proceeds through two phases. In the first phase, an initial clustering is obtained by the successive selection of representative objects until *k* representative objects have been found. The first representative object is the one for which the sum of the dissimilarities to all objects is as small as possible. This representative object is the most centrally located in the set of objects. Subsequently, at each step another object is selected. This object is the one which decreases the objective function as much as possible. In the second phase of the algorithm, it attempts to improve the set of representative objects and therefore also to improve clustering yielded by this set. In *k*-medoid cluster analysis centers are selected from the data set itself. Otherwise *k*-medoid is just like *k*-means algorithm. Typical aspects of these algorithms are that they provide a large number of statistics by which a thorough investigation of the clustering results is made possible, particularly by means of validation parameters.

To perform *k*-means cluster analysis on a data set. After choosing the number of clusters $1< c < N$ and initializing random cluster centers from the data set, the following steps were followed:

Step 1: From a data set of N points, *k*-means algorithm allocates each data point to one of c clusters to minimize the within-cluster sum of squares:

$$D_{ik}^2 = \left(x_k - v_i\right)^T \left(x_k - v_i\right); \; 1 \le i \le c; \; 1 \le k \le N, \quad (1)$$

where: D_{ik}^2 is the distance matrix between data points and the cluster centers; x_k is the k-th data point in cluster i; v_i is the mean for the data points over cluster i, called the cluster centers.

Step 2: Selecting points for a cluster which are having the minimal distances from the centroid.

Step 3: Calculating cluster centers:

$$v_i^{(l)} = \frac{\sum_{j=1}^{N_i} x_i}{N_i}; \quad (2)$$

$$\max\left|v^{(l)} - v^{(l-1)}\right| \ne 0, \quad (3)$$

where: N_i is the number of objects in the cluster i; j is the j-th cluster; l is the number of iterations.

In case of *k*-medoid clustering algorithms the Step 1 and Step 2 remain the same as *k*-means clustering. For *k*-medoid clustering at Step 3 the calculation for cluster centers is the same as *k*-means clustering even though it is called a fake cluster center. Step 4 in *k*-medoid clustering brings the difference in the two methods that the cluster center is chosen from the data points itself where as in case of *k*-means clustering this is not the necessary criterion.

Step 3: Calculating fake cluster centers:

$$v_i^{(l)^*} = \frac{\sum_{j=1}^{N_i} x_i}{N_i}. \quad (4)$$

Step 4: Choosing the nearest data point to be the cluster center:

$$D_{ik}^{2*} = \left(x_k - v_i^*\right)^T \left(x_k - v_i^*\right) \quad (5)$$

$$\text{and } x_i^* = \arg\min_i \left(D_{ik}^{2*}\right); \; v_i^{(l)} = x_i^*; \quad (6)$$

$$\text{until } \max\left|v^{(l)} - v^{(l-1)}\right| \ne 0, \quad (7)$$

where: D_{ik}^{2*} is the distance matrix between data points and the fake cluster centers; x_k is the k-th data point in cluster i; x_i^* is the fake data point in the cluster i which approximate to the calculated fake cluster center point and closer to any data point from the input data set; v_i^* is the fake mean for the data points over cluster i; $v_i^{(l)}$ is the center of cluster i after l number of iterations.

***Cluster Validation Measures*:** Cluster validity is concerned with checking the quality of clustering results. Determining the correct number of clusters in a data set has been the most common application of cluster validity (Bensaid *et al.* 1996). Different validity measures have been proposed in the literature, none of them is perfect by oneself, and therefore several parameters are used in this study, which is described below. These parameters are used for statistical cluster validation only and not for any model validation purposes.

***Partition Index (PI)*:** is the ratio of the sum of compactness to separation over all clusters. It is a sum of individual cluster validity measures normalized through division by the fuzzy cardinality of each cluster (Bensaid *et al.* 1996). *PI* is useful when comparing different partitions having equal number of clusters. A lower value of *PI* indicates better partition.

***Separation Index (SI)*:** On the contrary of partition index (*PI*), the separation index uses a minimum-distance separation for partition validity. Adopting the minimum-distance separation for partition validity is useful when searching for the 'right' number of clusters, but it does not useful when comparing different partitions having an equal number of clusters. This criterion favors the creation of a set of clusters that are maximally separated from one another (Xie, Beni 1991).

***Xie and Beni's Index (XB)*:** it aims to quantify the ratio of the total variation within clusters and the separation of clusters (Xie, Beni 1991).

Dunn's Index (DI): this index is originally proposed for the identification of 'compact and well separated clusters'. So, the result of the clustering has to be recalculated as it was a hard partition algorithm (Bezdek, Pal 1995, 1998).

***Silhouettes*:** A silhouette value s is expressed for each object as follows, Rousseuw (1987):

$$s = \frac{b-a}{\max\left(a,b\right)}, \qquad (8)$$

where: a is particular object i is in cluster A and a is equal to the average dissimilarity of i to all other objects in A. For every other cluster not equal to A, cluster B has the smallest average dissimilarity between its objects and i which is equal to b. The cluster B is the nearest neighbor of object i. Hence, silhouette value of an object measures how well an object has been classified by means of comparing its dissimilarity within an assigned cluster to its dissimilarity with its nearest neighbor.

3. Study Corridors and Data Collection

3.1. Study Corridors

The GIS layers of the study area for present analysis were obtained from a secondary source and were further enhanced in this study. A detailed road inventory survey was also carried out for preparing the digitized GIS base map of the road network. Five important road corridors of the city of Mumbai of Maharastra State, India are taken up for the present study. Grater Mumbai is an Island city with a linear pattern of transport network having predominant North-South commuter movements. Passengers move towards south for work trip in the morning hours and return back towards the north in the evening hours. Hence, four north-south corridors and one east-west corridor have been chosen for this study. Major roads like Eastern express highway extending up to south (Corridor-1), LBS Road extending up to south via Ambedkar road (Corridor-2), Western express highway extending up to marine drive (Corridor-3), SV road extending up to south via Veer Savarkar road (Corridor-4) and Versova- Andheri- Ghatkopar- Vashi (VAGV) (Corridor-5) are included. These five corridors overlapped on the GIS base map of Greater Mumbai are shown in Fig. 1.

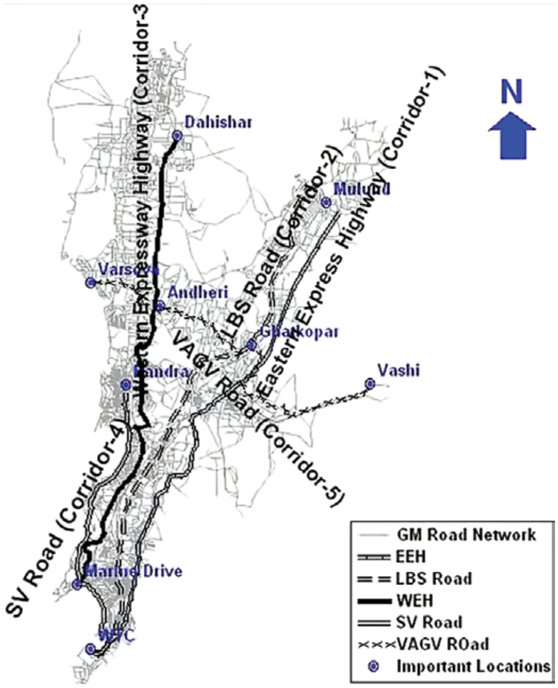

Fig. 1. Map showing selected corridors of greater Mumbai

These five road corridors, as a whole, cover 100 street segments. Segment is the directional stretch of a road section after a signalized intersection to the immediate next signalized intersection. These corridors were selected in preference from such a large road network which includes varying road geometric characteristics. Total length of road included in this study is approximately 140 kilometers. Roadway width varies significantly from location to location and in this study it includes two lane undivided roads to eight lane divided roads. Traffic movements are two-way on almost all segments except very few on which traffic movement is restricted to a single direction only. Direction-wise roadway capacity expressed in terms of passenger car unit per hour (PCU/hr) on these road corridors varies from 1500 to 3000. Similarly, free flow speeds (FFS) have a large variation on the road corridors i.e. on some segments FFS is 90 km/hr (kilometer per hour) and on some segments FFS is mere 25 km/hr. The travel speed also varies widely, which ranges from 5 km/hr on some road corridors with highly congested traffic during peak hours to 75 km/hr on corridors during off-peak hours. The traffic flow on these corridors is highly heterogeneous. There are significant percentage of two wheelers and three wheelers, this is one of the factors in retarding the average travel speed to as minimum as 5 km/hr. Some segments included in this study, however, are having very good flow characteristics with wide roads, footpath/shoulder, access facilities and separate lane for slow moving vehicles. In order to show the applicability of this study in other cities of India a similar survey was carried out in Kolkata City. Two corridors having varying geometric and surrounding environmental characteristics were taken into considerations i.e. one corridor was from Airport to Joka and the other corridor was from Airport to Ulberia. These two corridors are approximately 80 kilometer length; comprised of 50 street segments. The interesting fact on selecting these two cities for this study is that traffic composition and road geometric characteristics along with functionality brings the true variation that was required for this purpose. This wide variation in traffic characteristics along with the roadway geometrics makes these selected segments as a representative set for urban streets of Indian cities. Therefore, the methodology developed in this study for defining level of service criteria could be applicable to urban streets of Indian cities in general.

3.2. Data Collection

The probe vehicles used in this research work were mid-sized vehicles. These vehicles were fitted with Trimble Geo-XT GPS receiver, capable of logging speed data continuously at time intervals of one second. The GPS receiver provides both spatial and time/distance based data from which various traffic parameters were derived, including travel time, stopped time, travel speeds, and various congestion indices. In order to get unbiased data sets we used three mid-sized vehicles and took the help of three drivers on different days of the survey. Basically three types of data sets were collected. The first one is roadway inventory details, for which a data dictionary was prepared using Pathfinder office 3.0. Segment number, number of lanes on roadways, median type, parking conditions, pedestrian activity, road side development, access density, commercial activity and speed limits etc. were collected during inventory survey. All these help in classifying urban streets into number of classes. During the collection of inventory details proper segmentation technique was applied, which is after signalized intersection to after signalized intersection.

The second type of survey conducted was to find the free flow speeds on all these corridors. Before going for the free flow speed data collection, we need to know the time period during which traffic volume would be less than or equal to 200 vehicles per lane per hour. A detailed 24 hour traffic volume count survey was conducted by this group for another project in the month of April 2005. The traffic volume data were collected on 45 counting stations on seven screen lines covering the whole of Greater Mumbai region. Traffic volume per lane per hour was calculated for roads on these five corridors. From this data it was found that free flow traffic condition (less than 200 veh/ln/hr) is approaching at 12 mid-night and all road sections were having free flow traffic conditions from 1 AM to 5 AM. Hence, free flow speed on all these corridors was collected using GPS receiver fitted in probe vehicles during these hours. The third type of data colleted was congested travel speed. Congested travel speed survey was conducted during both peak and off-peak hours on both directions of travel on all the corridors. 10÷12 travel runs were made on all these five corridors consisting of 100 street segments. Collected data set were transferred to the office computer by using *Pathfinder office version* 3.00. The accuracy of the field data were made to improve significantly through the application of differential correction process. Differentially corrected files were visually checked before exporting those to a GIS database. Similar kind of survey was carried out in Kolkata also.

4. Results and Analysis

Directions-wise average free flow speeds on each segment were calculated by taking the average of all second-wise speed data collected. *K*-means cluster analysis with validation parameters like Partition Index, Separation Index, Xie and Beni's Index and Dunn's Index were applied to free flow speeds to find the optimal number of clusters. The optimal number of clusters helps in classifying urban streets into number of classes.

Results found from the validation parameters are plotted in Fig. 2. It was discussed that the lower value of Partition index indicates a better partition. From Figs 2a, 2b it is found that rate of decrease of Partition index and Separation index values beyond cluster number four is not very significant. Whereas, Xie and Beni's index reaches their local minimum at number of classes equals 5, hence justifies for five numbers of clusters as the optimal value. In Fig. 2d Dunn index value has reached local minima at clusters number 7. Classification of urban street into such large number of classes is undesirable. Comparing Dunn index values corresponding to number of clusters of 4, 5 and 6, it was found that local minimum was reached at cluster number 5. To find all such ambiguities before reaching the conclusion, it was decided to follow a middle path by selecting a number of clusters to be 5.

Once it was decided to classify urban streets into five classes, *k*-means clustering was applied on free flow speeds and free flow speed ranges were found for each street class I to V. Total 100 street segments were classified into five street classes based upon their free flow speeds. Direction-wise average travel speeds on street segments were calculated during both peak and off-peak hours under each street class. These data were used as input to *k*-means cluster analysis to find speed ranges of levels of service categories.

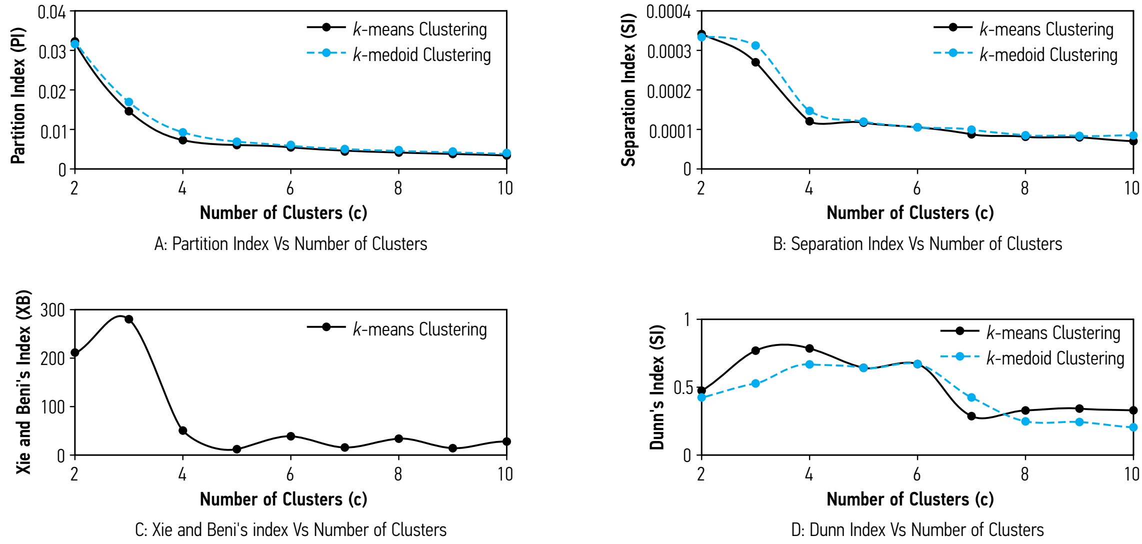

Fig. 2. Validation measures for optimal number of clusters

Comparing the speed ranges of street classes, it was found that speed ranges are not following the expected trend. Therefore, in order to get consistent results, as in Highway Capacity Manual (2000), in this study, urban streets were classified into four classes, which was further justified by two validation parameters i.e. partition index and separation index. Hence, firstly, *k*-means clustering was applied on free flow speeds and free flow speed ranges for each of the four street classes were found out. Secondly, *k*-means clustering was applied on average travel speeds to find out the speed ranges for each level of service categories. The free flow speed ranges of urban street classes and speed ranges of level of service categories using *k*-means clustering are shown in Table 1.

k-means clustering applied in defining level of service criteria of urban streets is discussed in previous paragraphs; in a similar fashion *k*-medoid clustering was also applied for defining level of service criteria. As discussed earlier, there lies a slight difference in algorithm details in these two methods of cluster analysis.

k-medoid clustering was applied both on free flow speeds and validation parameters to find the optimal number of clusters; and the output is also shown in Fig. 2. Considering to the criteria that partitions with lesser number of clusters are better, when the differences in the values of a validation index are minor for two consecutive numbers; Partition index and Separation index justifies for classification of urban streets into four classes although Dunn's index justifies in classifying urban streets into five classes.

The urban streets were classified into four classes. This classification is also satisfied by the geometric and surrounding environmental characteristics of street segments. Hence, free flow speed ranges were fixed for urban street classes. Speed ranges of level of service categories found from using *k*-medoid clustering on average travel speeds of urban street segments are shown in Fig. 3.

It may, however, bring into notice that this is a single variable (free flow speed) classification problem, both 'X' and 'Y' axes represent the average free flow speed, resulting in a 45 degree plot. Speed data under each level of service categories are shown by symbols, for example, in Fig. 3a data points that are shown by '+' symbol are falling under the same level of service category of 'A' in urban street class I. Silhouette values for level of service of urban street classes (I–IV) based on *k*-medoid clustering are shown in Fig. 4. In this plot, silhouette of levels of service is plotted in decreasing order. In order to obtain an overview, the silhouettes of all six levels of services are printed below each other. This way, the entire clustering was displayed by means of a single plot, which enabled us to distinguish a good cluster from weak ones. A wide silhouette indicates large silhouette values and hence a pronounced cluster. In this figure, speed points that lie well within their clusters (level of services) and the ones which are

Table 1. Speed ranges of LOS categories using *k*-means clustering

Urban street class	I	II	III	IV
Range of free-flow speed (FFS)	90 to 70 km/h	70 to 55 km/h	55 to 45 km/h	45 to 25 km/h
Typical FFS	75 km/h	60 km/h	50 km/h	40 km/h
LOS	Average travel speed (Km/h)			
A	>60	>49	>39	>33
B	>46÷60	>37÷49	>33÷39	>25÷33
C	>36÷46	>28÷37	>25÷33	>18÷25
D	>28÷36	>22÷28	>18÷25	>13÷18
E	>20÷28	>16÷22	>12÷18	>8.5÷13
F	≤20	≤16	≤12	≤8.5

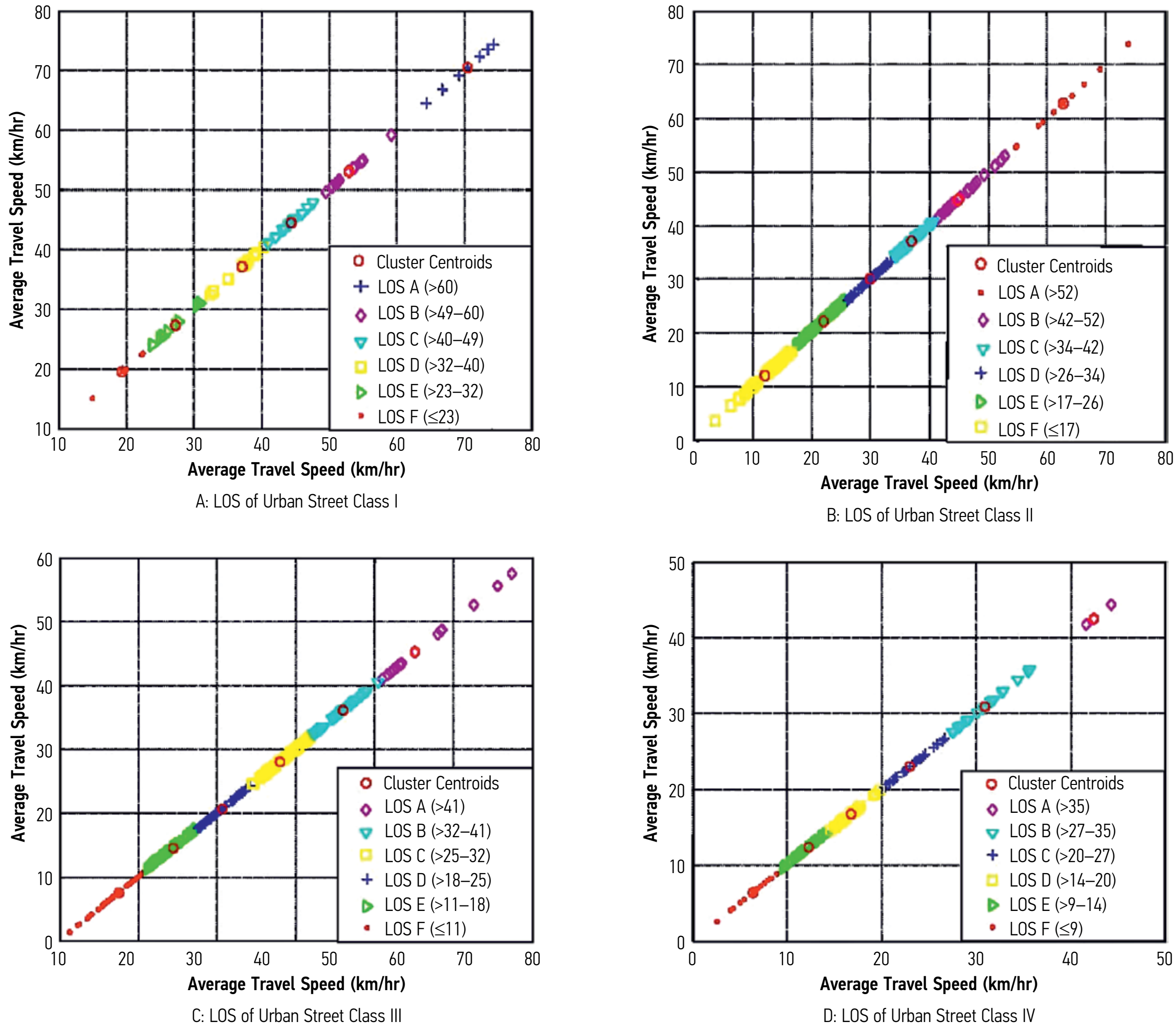

Fig. 3. *k*-medoid clustering of average travel speeds for LOS of urban street classes (I–IV)

merely somewhere in between clusters can be seen. The other dimension of silhouette is its height, which simply equals the number of objects in that cluster. It was found that for all four urban street classes, heights of level of service category D, E and F are comparatively higher than those heights of A, B and C categories.

This signifies that the probe vehicles traveled on these corridors more often at these poorer service levels. Speed ranges for levels of service categories using *k*-medoid clustering are shown in Table 2.

It can be compared that there lies significant difference in speed ranges for levels of service categories as shown in Table 1 and Table 2. Silhouettes validation parameter was used to select better clustering method between *k*-means and *k*-medoid methods for the classification of urban streets and level of service categories in urban Indian context. Silhouette width and silhouette coefficient of Urban Street Classes of both these clustering methods are shown in Table 3.

Average silhouette width for a street class is the average of silhouette values for all data points which lies within that class. And silhouette coefficient is the weighted average of silhouette widths of all groups formed out of the data set as a whole. Sum of average silhouette width for urban street classes (I–IV) using *k*-means clustering is equal to 2.473 (= 0.543 + 0.738 + 0.741 + 0.451). And sum of average silhouette width for urban street classes (I–IV) using *k*-medoid clustering is equal to 2.623 (= 0.651 + 0.672 + 0.777 + 0.523). Sum of silhouette coefficients for urban street classes (I–IV) using *k*-means clustering is equal to 2.893 (= 0.757 + 0.687 + 0.752 + 0.697). And sum of silhouette coefficients for urban street classes (I–IV) using *k*-medoid clustering is equal to 2.908 (= 0.737 + 0.722 + 0.716 + 0.733). Since average silhouette width and silhouette coefficients for urban street classes using *k*-medoid clustering are higher in values than *k*-means clustering values. Hence, *k*-medoid clustering method is the preferred and selected one in defining the level of service criteria of urban streets in Indian conditions.

In order to check the application of this level of service criteria; data collected in Kolkata city were tested. Free flow speed and average travel speed during both peak and off-peak hours on each of segments on both corridors were calculated. The street segments were classified into four classes based on free-flow speed, geometric and surrounding environmental characteristics. Also,

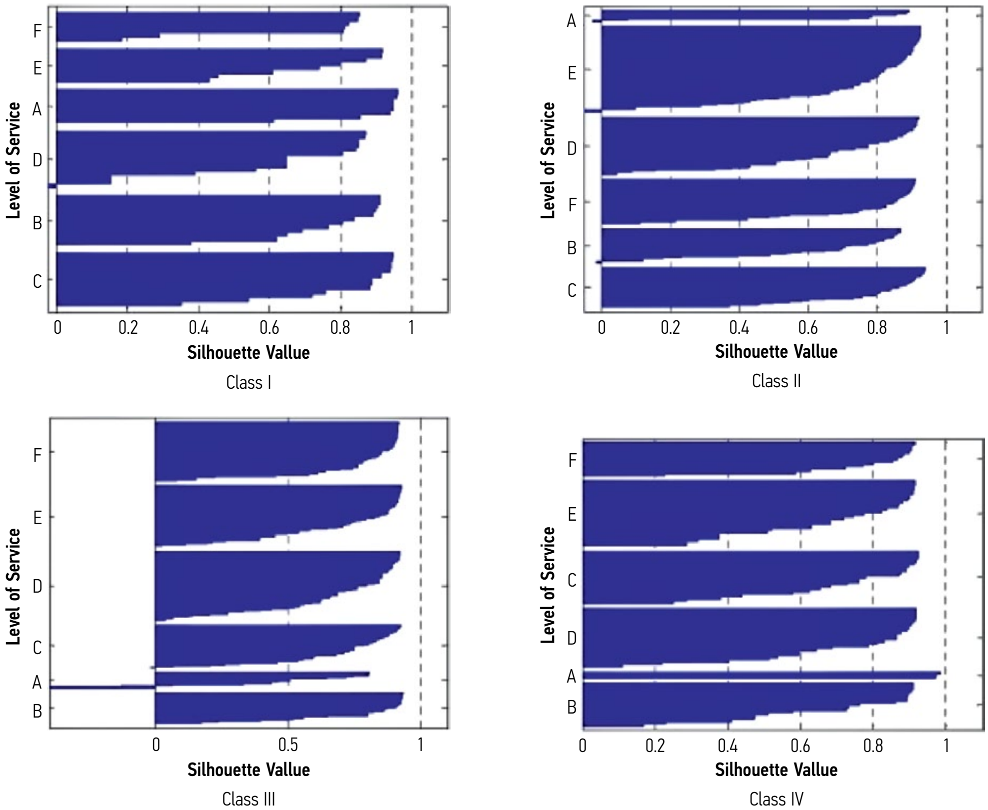

Fig. 4. Silhouettes plot for LOS categories of urban street classes (I–IV) using *k*-medoid clustering

Table 2. Speed ranges of LOS categories using *k*-medoid clustering

Urban Street Class	I	II	III	IV
Range of free-flow speed (FFS)	90 to 70 km/h	70 to 55 km/h	55 to 42 km/h	42 to 25 km/h
Typical FFS	75km/h	60km/h	50km/h	40 km/h
LOS	Average travel speed (Km/h)			
A	>60	>52	>41	>35
B	>49÷60	>42÷52	>32÷41	>27÷35
C	>40÷49	>34÷42	>25÷32	>20÷27
D	>32÷40	>26÷34	>18÷25	>14÷20
E	>23÷32	>17÷26	>11÷18	>9÷14
F	≤23	≤17	≤11	≤9

Table 3. Silhouette width and silhouette coefficient of urban street classes using *k*-means and *k*-medoid clustering

	Urban street class	Clustering methods	
		k-means	*k*-medoid
Average silhouette width (ASW)	I	0.543	0.651
	II	0.738	0.672
	III	0.741	0.777
	IV	0.451	0.523
Total ASW		**2.473**	**2.623**
Silhouette coefficient (SC)	I	0.757	0.737
	II	0.687	0.722
	III	0.752	0.716
	IV	0.697	0.733
Total SC		**2.893**	**2.908**
Total (ASW + SC)		**5.366**	**5.531**

levels of service provided by the street segments during peak and off peak hours were estimated. It is shown in Table 3. The percentage of travel runs under different levels of service categories found for urban street classes in Kolkata city during the survey period are shown in Table 4.

From Table 4 it has been noticed that the observed vehicle travelled at better quality of service under urban street class I, whereas under other urban street classes the observed vehicle travelled at medium quality of service during the observed period. Average travel speeds for the level of service categories expressed in percentage of free-flow speeds are calculated and the same is shown in Table 5. Values found using *k*-medoid method for the present context is compared with those values mentioned in Highway Capacity Manual (2000) and IRC 106… (1990). From this table it is found that there lies reasonable difference in the percentage values valid for Indian condition and those values mentioned in Highway Capacity Manual (2000) and IRC 106… (1990).

Table 4. Percentage of travel runs under levels of service categories for urban street classes in Kolkata city

Level of service	Urban street class			
	I	II	III	IV
A	52.78	0.00	3.92	15.56
B	30.56	22.22	11.76	24.44
C	8.33	22.22	21.57	22.22
D	2.78	22.22	31.37	17.78
E	2.78	27.78	19.61	6.67
F	2.78	5.56	11.76	13.33

Table 5. Comparison of percent FFS values for each LOS categories as obtained using *k*-medoid clustering method

Level of service	% FFS (HCM)	% FFS (IRC)	% FFS (*k*-medoid method)	Typical % FFS (*k*-medoid method)
A	90	90	80÷90	85
B	70	70	65÷80	75
C	50	50	55÷65	60
D	40	40	40÷55	45
E	33	33	30÷40	33
F	25÷33	25÷33	20÷30	20÷30

Notes: HCM – Highway Capacity Manual (2000);
IRC – Indian Road Congress (IRC 106… 1990)

5. Summary and Conclusion

In this study an attempt has been made in defining the level of services criteria for urban streets in the context of Indian cities. GPS was used to collect inventory and speed data on five major road corridors of Mumbai city, India. Collected data set was transferred to GIS build software called *TransCAD*. Two methods known as *k*-means and *k*-medoid clustering are described for the classification of urban streets and in defining speed ranges of levels of service categories. First of all, *k*-medoid cluster analysis was applied on free flow speeds and speed ranges of street classes were fixed. What is more, cluster analysis was applied on both peak and off-peak hour travel, speed data and speed ranges were found for the level of service categories by above mentioned methods. It was noted that no validation index is reliable only by itself that is why all the programmed indices were used. The optimum number of cluster using *k*-means clustering was identified to be five by the comparison of validation parameters. However, speed ranges for level of services under each street class did not follow the expected trend. Then, considering to the local condition and the classification adopted in Highway Capacity Manual (2000), the urban streets were classified into four street classes and speed ranges of LOS categories and the street classes were fixed. In a similar manner *k*-medoid clustering was applied and speed ranges in urban street classes and the levels of service categories were found out.

From this study, the following conclusions are made. Free flow speed ranges of urban street classes are different using *k*-means and *k*-medoid clustering methods. Free flow speed ranges of urban street classes mentioned in Highway Capacity Manual (2000) are significantly different from those found out in the present study. In particular, free flow speed range for urban street class IV mentioned in Highway Capacity Manual (2000) is 40 to 55 km/hr, whereas, in Indian context this range is found to be 25 to 45 km/hr and 25 to 42 km/hr using *k*-means and *k*-medoid clustering methods respectively. Speed ranges of the level of service categories found from using *k*-means and *k*-medoid methods are significantly different. Also, it was found that the urban street speed-ranges valid in Indian context are proportionately lower than the corresponding values mentioned in Highway Capacity Manual (2000). These lower values are highly heterogeneous traffic flow on urban roads with varying geometry. Average travel speed for each of the levels of service category expressed in terms of percentage of free flow speed is found to be different from those mentioned in Highway Capacity Manual (2000). Using validation parameter silhouettes a thorough investigation was carried out to find out the most suitable method, the one being applicable in Indian context. Silhouette width and silhouette coefficient values found for urban street classes are having higher values for *k*-medoid clustering than *k*-means clustering; hence, *k*-medoid clustering method is more applicable in defining the level of service criteria in Indian context. Silhouette height of good quality of service categories 'A', 'B' and 'C' is less than those heights of poorer service categories like 'D', 'E' and 'F'. This implies that probe vehicles traveled more frequently at poor service quality on urban streets in the present context.

References

Bensaid, A. M.; Hall, L. O.; Bezdek, J. C.; Clarke, L. P.; Silbiger, M. L.; Arrington, J. A.; Murtagh, R. F. 1996. Validity-guided (re)clustering with applications to image segmentation, *IEEE Transactions on Fuzzy Systems* 4(2): 112–123. http://dx.doi.org/10.1109/91.493905

Bezdek, J. C.; Pal, N. R. 1995. Cluster validation with generalized Dunn's indices, in *ANNES'95 Proceedings of the 2nd New Zealand Two-Stream International Conference on Artificial Neural Networks and Expert Systems*, 190–193.

Bezdek, J. C.; Pal, N. R. 1998. Some new indexes of cluster validity, *IEEE Transactions on Systems, Man, and Cybernetics – Part B: Cybernetics* 28(3): 301–315. http://dx.doi.org/10.1109/3477.678624

Highway Capacity Manual. 2000. Transportation Research Board, National Research Council. Washington, U.S.A. 1134 p.

IRC 106: Guidelines for Capacity of Urban Roads in Plain Areas. 1990. Indian Road Congress.

Jain, A. K.; Dubes, R. C. 1988. *Algorithms for Clustering Data*. Prentice Hall College Div. 304 p.

Lingras, P. 1995. Classifying highways: hierarchical grouping versus Kohonen neural networks, *Journal of Transportation Engineering* 121(4): 364–368.
http://dx.doi.org/10.1061/(ASCE)0733-947X(1995)121:4(364)

Lingras, P. 2001. Statistical and genetic algorithms classification of highways, *Journal of Transportation Engineering* 127(3): 237–243.
http://dx.doi.org/10.1061/(ASCE)0733-947X(2001)127:3(237)

Kaufman, L.; Rousseeuw, P. J. 1990. *Finding Groups in Data: An Introduction to Cluster Analysis*. Wiley-Interscience. 368 p.

Kim, K.; Yamashita, E. Y. 2007. Using a k-means clustering algorithm to examine patterns of pedestrian involved crashes in Honolulu, Hawaii, *Journal of Advanced Transportation* 41(1): 69–89. http://dx.doi.org/10.1002/atr.5670410106

Maitra, B.; Sikdar, P. K.; Dhingra, S. L. 1999. Modeling congestion on urban roads and assessing level of service, *Journal of Transportation Engineering* 125(6): 508–514.
http://dx.doi.org/10.1061/(ASCE)0733-947X(1999)125:6(508)

Marwah, B. R.; Singh, B. 2000. Level of service classification for urban heterogeneous traffic: a case study of Kanpur metropolis, in *Transportation Research Circular E-C018: 4th International Symposium on Highway Capacity: Proceedings*. 27 June – 1 July, 2000, Maui, Hawaii, 271–286.

Oh, C.; Ritchie, S. G. 2002. Real-time inductive-signature-based level of service for signalized intersections, *Transportation Research Record* 1802: 97–104.
http://dx.doi.org/10.3141/1802-12

Prassas, E. S.; Roess, R. P.; McShane, W. R. 1996. Cluster analysis as tool in traffic engineering, *Transportation Research Record* 1566: 39–48. http://dx.doi.org/10.3141/1566-05

Xie, X. L.; Beni, G. 1991. A validity measure for fuzzy clustering, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 13(8): 841–847. http://dx.doi.org/10.1109/34.85677